Panel CLARIN2021: The Role of Corpora for the Study of Language Use and Mental Health Conditions

Moderator: Henk van den Heuvel

CLARIN Annual Conference
Wednesday, 29 September 2021
10:15 - 11:00
Overview

Intro by Henk van den Heuvel

Two-minute presentations by panelists

• Gloria Galiardi
• Stefan Goetze
• Saturnino Luz
• Khiet Truong

Floor is open for Q&A

#CLARIN2021
Introduction Henk van den Heuvel

• Vibrant field: traces of mental health condition
• Multimodal: speech, text, video
• Vulnerable people: special category
• Collecting, processing and sharing these data
• How to avoid biases in software and tools
• Role of CLARIN
Introduction

2. ACA language resources

ACA language resources can be licensed only for research purposes. The end-user does not need to ask for usage permission but can access the resources via e.g. federated logs.

Examples of ACA licenses (and their CLARIN category labels)

Categories recommended by CLARIN with approved licenses

CLARIN ACA-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN ACA
CLARIN ACA-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN ACA
CLARIN ACA-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN ACA
CLARIN ACA-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN ACA
CLARIN ACA-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN ACA
CLARIN ACA-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN ACA

3. RES language resources

RES language resources have additional restrictions which require permission from the rights holder. These resources may contain material whose usage is limited due to copyright and/or personal data protection issues. In practice, these language resources require both using freedom logs to authenticate the end-user and sending a separate application to the rights holder for authorization, possibly including a research plan with the reasons.

Examples of RES licenses (and their CLARIN category labels)

Categories recommended by CLARIN with approved licenses

CLARIN RES-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN RES
CLARIN RES-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN RES
CLARIN RES-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN RES
CLARIN RES-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN RES
CLARIN RES-BY+NC+ND/NC/CC-LICENSED/CC-NV/CLARIN RES

https://www.clarin.eu/content/licenses-and-clarin-categories
http://delad.net
https://sshopencloud.eu/
RESEARCH TOPICS

• corpora of oral and written disordered productions
  → “digital linguistic biomarkers”
• lexical databases and language tests
  (to support Italian Speech and Language Therapist)

HOW COULD CLARIN FACILITATE FUTURE RESEARCH IN THIS FIELD?

• legal advice on data collection, processing, and storage according to EU regulation
• visibility and accessibility of language resources developed in/for the clinical domain
RESEARCH TOPICS

• Speech analysis and speech training systems for pathological speech
• Automatic speech recognition (ASR) for non-typical speech
• Signal assessment (intelligibility (SI), quality (SQ), listening effort (LE))
• Speech corpus collection - patients suffering from Morbus Parkinson or audiogenic speech disorders

HOW COULD CLARIN FACILITATE FUTURE RESEARCH IN THIS FIELD?

• More focus on speech, especially underrepresented recordings, (non-typical) speech
• Foster networking between researchers, clinicians and associations (e.g. formal and informal carers)
Saturnino Luz

Searching for cost-effective, non-intrusive digital biomarkers of neurodegeneration

- Speech and/or language?
- What kinds of data?
- On CLARIN: DementiaBank data
- Limitations:
  - Biases
  - Standardisation, comparability of results*
  - Data quality, pre-processing challenges (segmentation, diarisation, transcription, …)
- Attempts to address some of these limitations: ADReSS’20 and ADReSSo’21 Challenges at Interspeech
- How could CLARIN contribute:
  - Increase availability of ND speech and language data
  - Benchmarks, languages other than English
  - How to link to data on underlying clinical phenotype?
  - Privacy, data protection issues

(Computational) Paralinguistics, nonverbal aspects in speech communication, speech as a biomarker

- How to reach and collaborate with clinicians who have unused data?
- What do we know about the clinicians’ (or other “out-of-domain” researchers) needs for and use of SLP/NLP tools?
Discussion

• Use chat
• Raise hand
• Unmute yourself
Q&A